

# Logico-Linguistic Moleculism: Towards an Ontology of Collocations and Other Language Patterns

Dr Nikolay Milkov

“You shall know a word by the company it keeps.”

J. R. Firth

## Abstract

This paper is an exploration of the importance of the collocation approach in investigating language. This approach underlies a new conception of grammar that is: (i) intrinsically connected with lexis; (ii) investigates the language as it is naturally used in life; (iii) can be developed as a corpus-driven grammar.

The collocation approach in language investigation is also examined from the perspective of some recent developments in the philosophy of language. The conclusion reached is the identity between philosophical ontology, linguistic ontology and computational ontology.

## 0. Opening

In this paper I will try to assess achievements in lexicography over the last few decades from the perspective of some recent developments in philosophy of language. One of the conclusions arrived at is *prima facie* astonishing: there is no difference between philosophical ontology and computational ontology. Here some elucidation is in order. Usually, philosophical ontology is understood as a classification of entities. In contrast, the ontology of information-processing is interested in representations in people’s head. It does not pay attention to the truth but to the computational tractability of its concepts.

In short, it is presumed that “there is a certain tension between the philosophical understanding of the term ‘ontology’ and the use of the term that is proliferating through the information-processing world” [22], p. 3.

There is no such a tension in the concept of language developed in this article. To be more precise, in connecting situations closely with the words which depict them through the

relation of fitting, we assume one ontology instead of two, or three—linguistic, computational and philosophical. In what follows, supportive evidence for this thesis will be offered.

## **1. Mainstream Philosophy of Language and my Approach**

Traditionally, the mainstream philosophers of language ignore lexicography. A typical example of this is the most influential philosopher of language from the second half of the twentieth century, Willard Quine, who is still convinced that “nothing, apparently, could be more remote than meanings from lexicographical concern” [19], p. 58.

Following some points of J. L. Austin’s philosophy of language, Quine’s pupil Donald Davidson accepted that a proper theory of meaning must explain the complex skill which is constitutive for the mastery of a language. Theory of meaning should be nothing but knowledge of how we use language.

In this paper I develop this point further, making it a radical revision. My rationale for doing so is that Davidson’s understanding of language-use is too general and thus too abstract. In contrast, I am going to advance a theory of how we creatively speak/write language. The motivation for this emerges from the realisation that the skills required for active language-producing are much more complicated than the skills required for language-apprehension. We are undoubtedly all aware from personal experience that to speak and write so as to convince an audience requires a higher level of ability than that needed to understand what other people say or have written.

### **1.1. Criticism of the Chemical Approach to Language**

Wittgenstein’s *Philosophical Investigations* begin with a criticism of the good old belief, embraced, for example, by St. Augustine, that little children learn their native language by learning words. In contrast, Wittgenstein accepts (following Gottlob Frege) that children learn natural languages through learning sentences. Moreover, he maintains that single words have no meaning. Only sentences have meaning. In contemporary philosophy of language, this understanding is called the *context principle*, p. words only have meaning in the context of a sentence.

An example: The most influential philosopher of language in Britain in the last three decades, Michael Dummett, who in many respects follows the already cited Donald Davidson, accepts that the atoms of language are words, while its molecules are sentences. Since the elementary units of discourse are sentences, the central theme in his philosophy of language is “logical molecuism”, as opposed to Bertrand Russell’s “logical atomism”.

Contrary to this understanding, I would argue that the mere chemical division of language-units into words and sentences—atoms and molecules—is not sufficient to explain how language really functions. Rather, I propose the biological model, in terms of which language is more like a living organism, and so is to be divided into words (atoms), collocations (molecules), sentences (cells), paragraphs (organs), stories (living individuals). Only by accepting this model can we reach a plausible theory of creating language-units.

## 2. What Are Collocations?

In addition, I would maintain that the backbone of language is in fact its molecules. These are what in recent years have been called *collocations*, and earlier were called “locutions”, “phrases”, “idioms”, “patterns”, “clusters”, “lexical phrases”, “composites”, “gambits”, “routine formulae”, “phrasems”, “prefabricated routines and patterns”, “formulaic language”, etc. (see [11], p. 7).

In short, collocations are recurrent combinations of words. According to the definition of Cowie, Makin and McCaig, it is characteristic of collocations that “the sense of the whole [of the collocation] cannot be arrived at from a prior understanding [of the words] of the parts” [3], ii. p. x.

The term *collocations* was first intensively discussed in linguistics in the early 1950s by J. R. Firth (see [5], pp. 194 ff.). As this author has noted elsewhere: “Logicians continue to treat words and sentences as if they somehow could have meanings in and by themselves. . . . Both these pre-suppositions are misleading in linguistics and wholly inadequate for the handling of speech events which is the main objective of the discipline.”<sup>1</sup> [6], pp. 12–13.

The whole idea of this approach to language is that words, plus grammar, are insufficient both for learning a language, as well as for its producing in a convincing shape; the collocations of words are at least as important. Why is this? As it was recently put, it is not difficult to notice that “natural language has a certain block-like character. Words tend to occur in the same clusters again and again. When we speak or write it is therefore often more apposite to say that we move from one cluster to the next than to say that we move from one word to the next.” [13], p. vii. Take as examples these three words:

- **panic** – We can *cause* or *create* it; or to *avert* or *prevent* it; it can *spread*. These are the five verbs that are usually connected with this word. When we say that panic + so-

---

<sup>1</sup> I shall return to the point that speech events are the main ingredient in language ontology in § 3.

and-so (verb), we are urged through the language-convention to use one of these five verbs. The use of some other verb is risky; it can be a sign of a very good style; or that we are using the language incorrectly.

- **map** – We can *draw*, or *trace* a map, or *read* a map. These are all the conventional verbs used with this word.
- **problem** – We can *cause*, *create*, *pose*, or *present* a problem; or we can *address*, *tackle*; *bring up*, *raise*; *resolve*, *settle*, *solve* a problem.<sup>2</sup>

What is rather surprising and counter-intuitive here is that every word collocates with a very small number of other words. It is counter-intuitive since we believe that the collocability of words is almost unlimited; that collocations result from permutations of words. This conviction explains, among other things, why the collocation approach to language was neglected for so many years. Actually, collocations are astonishingly small in number and so tractable. As cleverly noted some 25 years ago, “our psychological lexicon contains large numbers of multiple-word units [collocations]—stock phrases of various sorts. Indeed, there are probably at least as many of these as of single words.” [12], p. 65 n. 5. But not significantly more. This was recently demonstrated by the first comprehensive dictionary of collocations [1], which is astonishingly small in volume.

To put all this into three words: words like company—but not much company. This understanding has also been termed the *idiom principle*. “The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.” [21], p. 110.

## 2.1. The Underlying Philosophy of Language

The leading idea of the collocation approach to language is that language functions quasi-mechanically, following possibilities already traced—through the language practice—for combining words into strictly fixed clusters. The creative use of language is just as mechanical as any other craft (or engineering) is. This, as we are going to see in what follows, makes the task of computer-processing with language-units pleasant and fruitful.

It is true that the idiom principle “constitute[s] a single language choice rather than a series of choices” [11], pp. 230–1. Nevertheless, there is a room for creativity here (and we are

---

<sup>2</sup> All three examples are from [1].

going to see this in § 3.3). This creativity is the same in form as that of inventing a device, or a new engine: the purpose is to construct a new pattern, a new paradigm, in order to use it in a mechanical way for long period of time.

The collocative meaning is an idiosyncratic property of the individual words. In fact, every word has a conventionally limited *range* that determines with which other words it can collocate (see [2], pp. 52 f.). For instance, *putrid*, *rancid* and *addled* are synonymous; nevertheless, *putrid* collocates with fish, *rancid* with butter, oil, lard, and *addled* with eggs.

Every natural language can be seen as the sum aggregate of the words in all their ranges. Here we can set out our first ontology (**O1**), which has a purely linguistic character. I will base this ontology on the idiosyncratic logic I develop in [16]. According to **O1**, every word can be seen as a spatial *object*, a polyhedron that has many different faces—these are its nuances, its different meanings. It can be activated, in order to join the “company” of other words, on any one of its sides. Both phrases (collocations) and sentences are composed by way of connecting different sides of the words in strings, or formulas. Most of them are continuous strings. Of course, there are also discontinuous collocations, by which an alien word enters in. For example: *to set* [SOMETHING] *in motion*. This, however, hardly changes their nature as continuants.

The whole of the language is the sum of all possible collocations of its words, which are now implicit in it. One particular collocation is an actualisation from this sum.

### **3.How Do Collocations Function?**

Through the centuries of its functioning, language usage has strictly determined the valid collocations of the natural language. This means that every event, every situation, can be expressed through its (one) appropriate collocation. When we speak, or write, or translate, we put our intuition to work: we consider and choose the right expression (word). We are, more precisely, looking for the collocation that is most appropriate for this life-situation.

As already noted, active language-production is nothing but a craft. This craft is rooted in the purely mechanical activity of finding, among the clusters of words which are in usage, those one which are most appropriate for exactly this situation. This point explains the following practice set out by one of the most popular translators of philosophical literature in German (from English and Italian): “Just as we can tell the tailor where he went wrong, so we can point out to the speaker who has chosen the wrong word that his description did not fit the subject of his portrait.” [20], p. 142. Wittgenstein has put this point so: one “chooses between words as between similar but not identical pictures” [27], § 54.

### 3.1. Philosophical Ontology of Language

Evidently, the successful producing of language is an activity similar to that of tailoring. But how is this activity possible?

In answering this question we follow Wittgenstein's argument from his famous *Tractatus logico-philosophicus*, according to which language constructs pictures (models), as if in experiment, with the help of which we replicate facts of reality [26], 4.0311. When we speak, or write, we construct a second reality with the aim of expressing ourselves in communication; every meaningful sentence or proposition constructs such fake realities (models). In this way, language-producing follows a strategy of "world constructing", or an *Aufbau*-programme.

An important aspect of this programme is the Principle of Representing, according to which if two formations have the same form (to use Wittgenstein's term here, the same "logical multiplicity"), then the first can represent the second, and vice versa—the second can represent the first. This explains why we can completely understand a situation on seeing a black-and-white photograph, which surely is far from being a perfect picture of the situation.

Language functions similarly. (i) Its words picture objects of the real world in an exact way—by referring to them. We can thus use them to represent real objects in a most precise manner. Of course, a dictionary consisting of such terms is a poor one—it sets out a static world, a world of objects that do not interact or relate to one another. A typical example of such types of dictionaries are the Picture Dictionaries, the primary objective of which is to make their readers familiar with the names of different objects.

(ii) Real life, however, is a world of facts, events and processes, not of static objects. Life cannot be depicted by a Picture Dictionary. This is better done by Dictionary of Collocations; that alone can deliver life's most variegated situations and events.

In this second onslaught aimed at clearing up the linguistic ontology (**O2**), we accept that language is nothing but an interweaving of words, through which we build up, as if for experiment, possible worlds, in the way described above. Now, even one word determines a world. Clear as this world is, however, it is characteristically poor and unstable—we are not sure how to connect it with the other worlds our language produces. Matters are altogether different with the worlds constructed by way of collocations. Metaphorically, this is so since they pin new meanings down by means of two ontological strings (meanings), thus rooting them in the soil of our language practice.

### 3.2. Linguistic Phenomenology

J. R. Firth was first to note, in the early 1950s, that “any text can be regarded as a constituent of a *context of situation*” [6], p. 175. Roughly at the same time,<sup>3</sup> the theory of collocations was advanced in a philosophy of language that was called linguistic phenomenology. It was first developed, in the late 1940s and in the 1950s, by the British philosopher J. L. Austin (1911–1960). One of its objectives was to thoroughly reveal the possible semantic of the word under scrutiny. The means for this is the listing of all the *situations* in which it can occur, the most extravagant situations included. Extravagant situations, after all, bring to light nuances of the word’s meaning which we are inclined to neglect.

This procedure is nothing but examining a word—finding out its place in the system of words, in the net of concepts. Every word has its own specific place in this net. The underlying understanding is that from the point of view of the *philosophical ontology* (O3), for every situation there is exactly one word or phrase that connotes it in a precise way. It cannot be replaced by another word (phrase, collocation). This is a consequence of the fact that in their long history, natural languages have made the words distinct enough in order to fit every possible situation. What the theory of collocation added to this perception of Austin’s is that with the help of collocations, the resources of language to cover every possible situation are considerably and decisively augmented.

Unfortunately, despite the fact that the meaning of the words is exactly fixed, it often happens that words are used one-sidedly, and so incorrectly. This characteristically occurs in ideologically-laden disciplines such as philosophy, theology, history, sociology. The task of the linguistic phenomenologist is to bring us closer to the correct meaning of the words. In this sense, she/he rectifies the natural language.

### 3.3. Metaphors and Poetry

According to this apprehension of how language functions, good style both follows the cannon set out through the usage, and at the same time prepares the ground for a step-by-step creation of a new usage.

Indeed, the good literary stylist, the poet, for example, uses well-established models, collocations, which are the most appropriate (fitting) phrases available that are required for depicting the situation (imaginative, or real) which he intends to communicate. At the same time, the poet, or the orator, the language stylist, also introduces new collocations (models),

---

<sup>3</sup> Here we will not discuss the question whether Firth and his friends influenced British philosophy of language of the time, or whether the direction of influence was in fact a converse one.

thus discovering new dimensions of her/his language. As a rule, the new collocations are made on analogy with already existing ones. The mastery here consists in finding out of new similes between the words, thus connecting them in new language-molecules.

This understanding also suggests a convincing explanation of how metaphors function. Indeed, as Donald Davidson has shown, metaphors “mean” in exactly the same way in which all other words do (see [4], p. 245). What makes metaphors different is that they are characteristically ambiguous, thus stimulating our insight and imagination to fabricate ever new collocations with them. My theory of language suggests that this ambiguity arises with the fact that metaphors’ facets (or nuances) are radically more in number than those of plain words.

My understanding of how language functions also makes the method of hermeneutics, which strives to explain how the words of language are connected to one another, more comprehensive. According to the one of the most celebrated hermeneuticists, Martin Heidegger, the relation between a word which is already there, and the word that we now must find and put next to it, in a collocation, is “a poetical relation”. The first one “calls” the second. The connection between the words in a collocation is realised through listening out for it, following its call: “Das Nennen verteilt nicht Titel, verwendet nicht Wörter, sondern ruft ins Wort. Das Nennen ruft. Das Rufen bringt sein Gerufenes näher.” [9], p. 18.

#### **4.A General Assessment of the Collocation Approach to Language**

My theoretical approach is similar to that of intuitionism, which was introduced into the philosophy of mathematics by L. E. J. Brouwer at the beginning of the twentieth century. Brouwer refused to accept mathematical entities which are not constructed *de facto*. He opposed those abstract notions that are defined through our intuition with the phrase “and so on”.

Similarly, my thesis is that when a person tries to deduce his English as a foreign language from the conventional or analytical grammar that he has learned by heart, plus the words (lexicon) he has learned, he is doomed to failure. In contrast, I maintain that a foreign language can be successfully learned with the purpose of using it creatively, only step by step, in an effort to master all the possible combinations that the words which are going to be used in communication can have. This task is not impractical.

Indeed, this conception can be seen as a matter of fact, if we bear in mind our second ontology of language (**O2**). Remember that according to this, language-production is nothing but constructing models of reality. From this perspective, it is clear that language has charac-



teristics close to those of reality—for example, its nature is, among other things, topographical. So in a sense, learning language is similar to getting acquainted with an environment, with a country in which we must act. This explains why there are no effective methods of learning it extremely fast and intensively. Every newcomer should travel through the land, picking up all its particulars separately.

To sum up, my approach has three primary characteristics. It is based: (i) on the usage, not on analysis; (ii) on lexis, not on grammar; (iii) on language teaching. It describes patterns, not structure.<sup>4</sup> It can be generally characterised as a dislike for theories that lead to a refusal to accept a grammar in the conventional sense.

#### **4.1. History of this Approach**

Only in recent years has there been evidence of the birth of the new approach to language, which J. L. Austin had once dreamed of. As might be expected, it was developed by lexicographers, not by philosophers. Here is a short history of the theory of collocations in connection with language acquisition.

(i) The approach was tentatively developed by Harold Palmer in his *A Grammar of English Words*, Longman, 1938. This author was the first to realise that what matters in language-learning is not only the abstract grammar of forms, but also the concrete grammar of words. This means nothing less than that every word has its own grammar which is to be specified or learned separately.

(ii) As already noted, the term “collocation” was first extensively used by J. R. Firth in [5], pp. 194 ff.

(iii) The idea of constructing a dictionary of collocations was put forward as early as 1974 in the article “Dictionary” of *Encyclopaedia Britannica* (vol. 5, p. 721b) which says: “‘Dictionary of collocations’ may be a step forward in future lexicography.”

(iv) J. R. Firth’s idea of new, concrete grammar of words was followed by his colleague Ronald Mackin, who worked on the project in the 1960s, and later, in 1975–83, together with his friends A. P. Cowie and I. R. McCaig, published the *Oxford Dictionary of Current Idiomatic English* (2nd ed. 1993) [3].

(v) In 1986 Morton Benson, Evelyn Benson and Robert Ilson published the *Combinatory Dictionary of English* (2nd ed. 1993) [1], which is quite evidently based on collocations.

---

<sup>4</sup> This point was first articulated in [10].

Despite the good progress made by the project for a new approach in both lexis and grammar, all these authors refused to call their products a Dictionary of Collocations. R. Mackin, A. P. Cowie and I. R. McCaig called the two volumes of their work [3], respectively, *Dictionary of Phrasal Verbs*, and *Dictionary of English Idioms*. In Germany, a work compatible with this one, written by Günter Drosdowski, was called *Stilwörterbuch der deutschen Sprache* (7. Auflage, Mannheim: Dudenverlag, 1988).

(vi) In 1994 Göran Kjellmer published *A Dictionary of English Collocations*, 3 vols., Oxford: Clarendon Press [13].

(vii) In recent years, the collocation approach to language was realised in the fields of the grammatical collocations (more on these in section 5.3) in *COBUILD: The grammar pattern series* (see [7], [8]).

## 5. Formal Matters

In this section I am going to track down the *logic of collocations*. We need this in order to make the transition in the next section (§ 6) to the final theme of this investigation: outlining the shape proper dictionaries of collocations and other linguistic patterns ought to have, using a corpus-driven grammar.

### 5.1. Idioms as Collocations

There are no hard and fast distinctions between idioms, collocations proper and free combinations of words: collocations constitute a continuous universe. Here is their loose classification [3], ii. pp. xii–xiii:

- *pure idioms*, or idioms in the strict sense, such as *miss the boat*, or *kick the bucket*.
- *figurative idioms*; these scarcely permit variations, or pronoun substitution. Once formed, they do not change. Examples: *beat one's breast*, *burn one's boats*, etc.
- restricted collocations, sometimes referred to as *semi-idioms*; in these, one word has a figurative sense, whereas the other has a literal sense. Example: *jog one's/somebody's memory*.
- open collocations, or *free idioms*; in these both elements (words) have their literal meaning, and are also freely re-combinable. Thus I can say *fill the sink*, but also *drain the basin*.

Despite the considerable difference between these kinds of idioms, from a logical point of view, they are all collocations, in the sense that they consist of two meanings (words) combined into one, in a way that produces a new meaning. In short, idioms can be seen as a type of collocation. Indeed, both pure idioms and restricted collocations are built upon the same logical principle—they are conjunctions of two words—meanings.

## 5.2. Criteria for Collocationhood

Different collocations have various degrees of cohesion, or collocability. Thus, the pure idioms are monolithic expressions, the compounding words of which are quasi petrified or congealed (see [14]). Some authors speak about an ‘organic interrelationship’ between the words in them.<sup>5</sup> The free idioms are another kettle of fish.

This difference leads me to introduce the notion of the *collocationhood* of the collocations in general. There are three different criteria for collocationhood.

(i) Some believe that the only requirement for collocability is the frequency with which two words are used together. This, however, is scarcely the case. If we analyse a sufficiently large corpus of English, we will see that some word-combinations occur often, despite the fact that there is no connection in sense between them, for example, *day but*, or *however in the*.

(ii) The recurrent sequences of words need to be grammatically well-formed—they would have to conform to a pattern established through linguistic practice (see [13], p. xiv) and they can deviate from it to a small degree only. And this is where *creativity* comes in when introducing new language-patterns. On the one hand, the new collocation that the stylist introduces should be stimulating; on the other hand, it shouldn’t be unintelligible. The use of unintelligible collocations, rather, is a mark of the speech of a person who is not a native speaker, and who hasn’t mastered the language well enough.

(iii) Besides being grammatically well-formed, collocations are also to be native-like; with their help, the native speaker should be in a position to routinely convey his meaning (see [18], p. 191). Example: It is true “that among the selectional restrictions of *drink* there would be a restriction specifying that the object of *drink* has the feature [LIQUID]. Thus *drink tea*, *drink water*, *drink beer* all conform to this rule, as do *drink soup*, or *drink medicine*, but

---

<sup>5</sup> At the beginning of this paper (§ 1.1) I have already used the biological metaphor to describe my approach to language.

it is only those in the first group, all commonly used, that should be called collocations, since those in the second hardly ever occur.” [13], p. xviii.

### 5.3. Grammatical Collocations. Colligations

Besides purely lexical collocations, there are also grammatical collocations. The discrimination between the two was first made by J. R. Firth. In his later articles, he expressed it this way. When dealing with lexis, he spoke of collocations; when he dealt with the linking of the chunks of language that are grammatically well-formed, he spoke of *colligations* instead (see [6], p. 181).<sup>6</sup>

Today the difference between collocations and colligations is marked this way. When speaking of syntax, linguistic ontologists use the term “patterns”; when speaking of lexis, they use the term “collocations”. The pattern approach in grammar was approbated, for example, in [7] and [8].

**Sentences as Collocations:** By the same method of sticking together two chunks of language into one in a uniform manner, we can easily present the making up of sentences. As a matter of fact, this job is rather technical. In some sense, sentence-formation can be apprehended as the building up of collocations of collocations,<sup>7</sup> or as colligations of collocations: as an extensional connection of two or more chunks of language into one sentence—of two molecules in one cell (sentence). The conclusion is that sentences too are constructed by way of connecting of two or more linguistic chunks into a linguistic aggregate.

This understanding has led its defenders to assume that “there is a strong tendency for sense and syntax to be associated. The correspondences [between them] are overwhelming” [21], p. 65. It finally became evident that the old distinction between syntax and semantics is not so important as was formerly believed.

### 5.4. The Logic of Collocations

Collocations—which can, for convenience, be split up into grammatical and lexical ones—can easily be formalised. For example, this way:

---

<sup>6</sup> Incidentally, the term *colligations* reveals the sources of Firth’s ontological philosophy of language. Here it must be remembered that this term was first used in W. H. Walsh’s philosophy of history, in the sense that historical events are ordered (colligate) in relation to one another, thus composing complex wholes. The events do not cause one another, but nest in, and permeate in one another (see [25]: 59 ff.)

<sup>7</sup> And indeed the dictionaries of phrases are often presented as dictionaries of sentences (see, for example, [17]).

(i) Grammatical collocations consist of: noun + prepositional combination (G1); noun + to + infinitive (G2); grammatical construction + clause (G3); preposition + noun combination (G4).

(ii) Lexical collocations consist of: verb + noun (L1); adjective + noun (L2) (see [15]); noun + verb (L3); noun + noun (L4); adverb + adjective (L5); adverb + verb (L6) (see [1], pp. xv–xxxiv).

A similar formalisation can also be elaborated for sentences. Such a project was developed, for example, by P. F. Strawson, who first advanced a programme for analysing sentences through type-differences other than the formal-grammatical ones. In his *Individuals* [23], Part Two, but especially in the little-discussed *Subject and Predicate in Logic and Grammar* [24], Strawson replaced the traditional “logical form” view of sentences and propositions with recurrent “patterns” view, accepting that a meaningful sentence is a result of a correct connection of two types, or “categories”.

## 6. Computational Ontology of Language

But what is the connection of the collocation (colligation) approach in investigating language with the computational ontology of language? We see it, above all, in the projects for comprehensive dictionaries of collocations/colligations, based on *corpus-driven linguistics*. The latter follows “a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and calculates” [11], p. 15. The task requires a full machine-supported coverage of all words in all their meaningful combinations.

In corpus-driven linguistics, the authentic linguistic data have priority, not some theory which is extracted from the real language. The linguist following this approach does not organise the material for theoretical reasons, because it seems interesting to her/him. Rather, his/her data are organised by means of a software program. This approach is the opposite of the *corpus-based solution*, the problems of which do not arise as an outcome of corpus research, but from language-research that is driven by the intuition of the linguist.

The composition of collocation (colligation) dictionaries based a corpus-driven approach is, of course, a difficult job. The best achievement in this realm so far, the three-volume *Dictionary of Collocations* [13], is based on a corpus of only one million words (the so-called *Brown Corpus*). For this reason, as the author himself notes, it falls short of being comprehensive.

Fortunately, in the last few years the preparation of more compendious collocation dictionaries has become a more feasible task. For this purpose the *Bank of English* has been established, and now comprises some 300 million words.

From the perspective of our investigation, we see the envisaged programme for comprehensive dictionaries of collocations (colligations) as based on a computational ontology of language (**O4**), which is only an aspect of the logico-linguist molecuism—besides its three other aspects, already discussed in the paper.

## References

1. Benson, M., *et al.* (1993), *The BBI Combinatory Dictionary of English. A Guide to Word Combination*, Amsterdam: John Benjamins, 1st ed. 1986.
2. Carter, R. (1987), *Vocabulary: Applied Linguistic Perspectives*. London: Hyman.
3. Cowie, A. P., Mackin, R., and McCaig, I. R. (1993), *Oxford Dictionary of Current Idiomatic English*, 2 vols., 2nd ed., Oxford: Oxford University Press.
4. Davidson, D. (1984), *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.
5. Firth, J. R. (1957), *Papers in Linguistic*, Oxford: Oxford University Press.
6. Firth, J. R. (1968), *Selected Papers of J. R. Firth. 1953–59*, ed. by F. R. Palmer, London: Longmans.
7. Francis, G., Hunston, S., and Manning, E. (1996), *Collins COBUILD Grammar Patterns 1: Verbs*, London: HarperCollins.
8. Francis, G., Hunston, S., and Manning, E. (1998), *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*, London: HarperCollins.
9. Heidegger, M. (1959), *Unterwegs zur Sprache*, Pfullingen: Neske.
10. Hornby, A. S. (1954), *A Guide to Patterns and Usage in English*, Oxford: Oxford University Press.
11. Hunston, S., and Francis, G. (2000), *Pattern Grammar*, Amsterdam: John Benjamins.
12. Kelly, E. F., and Stone, P. J. (1975), *Computer Recognition of English Word Senses*, Amsterdam: North-Holland.
13. Kjellmer, G. (1994), *A Dictionary of English Collocations*, 3 vols. Oxford: Clarendon Press.
14. Leach, G. (1981), *Semantics*, 2nd edn., Harmondsworth: Penguin.

15. Mikhail, E. H. (1994), *The Cassell Dictionary of Appropriate Adjectives*, London: Cassell.
16. Milkov, N. M. (1999), "Tractarian Topology. A New Interpretation of the Objects in Wittgenstein's *Tractatus*", *Contributions of the Austrian Ludwig Wittgenstein Society*, 7, ii. pp. 53–8.
17. Pafen, K.-A. (1980), *Deutsch-russisches Satzlexikon*, 2 Bände, Leipzig: Enzyklopädie, 1. Ausgabe 1966.
18. Pawley, A., and Syder, F. H. (1983), "Two Puzzles for Linguistic Theory: Native-like Selection and Nativelike Fluency", in Richards, J. C., *et al.* (eds.), *Language and Cognition*, London: Longman.
19. Quine, W. V. O. (1990), *Pursuit of Truth*, Cambridge (Mass.): Harvard University Press.
20. Schulte, J. (1988), "Remarks on Sprachgefühl", in C. Nyíri, and B. Smith (eds.), *Practical Knowledge: Outlines of a Theory of Traditions and Skills*, London: Croom Helm, pp. 136–46.
21. Sinclair, J. M. (1991), *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
22. Smith, B. (2000), "Ontology: Philosophical and Computational", <http://wings.buffalo.edu/philosophy/faculty/smith/articles/ontologies.htm>
23. Strawson, P. F. (1959), *Individuals*, London: Methuen.
24. Strawson, P. F. (1974), *Subject and Predicate in Logic and Grammar*, London: Methuen.
25. Walsh, W. H. (1951), *Introduction to Philosophy of History*, London: Hutchinson.
26. Wittgenstein, L. (1922), *Tractatus Logico-philosophicus*, London: Routledge.
27. Wittgenstein, L. (1953), *Philosophical Investigations*, Oxford: Blackwell.